

mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer

Year: 2020 | Citations: 2865 | Authors: Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou

Abstract

The recent “Text-to-Text Transfer Transformer” (T5) leveraged a unified text-to-text format and scale to attain state-of-the-art results on a wide variety of English-language NLP tasks. In this paper, we introduce mT5, a multilingual variant of T5 that was pre-trained on a new Common Crawl-based dataset covering 101 languages. We detail the design and modified training of mT5 and demonstrate its state-of-the-art performance on many multilingual benchmarks. We also describe a simple technique to prevent “accidental translation” in the zero-shot setting, where a generative model chooses to (partially) translate its prediction into the wrong language. All of the code and model checkpoints used in this work are publicly available.